

NPInter v2.0: an updated database of ncRNA interactions

Jiao Yuan^{1,2}, Wei Wu^{1,2}, Chaoyong Xie^{2,3}, Guoguang Zhao³, Yi Zhao^{3,*} and Runsheng Chen^{1,*}

¹Laboratory of Noncoding RNA, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China,

²University of Chinese Academy of Sciences, Beijing 100049, China and ³Bioinformatics Research Group, Advanced Computing Research Laboratory, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China

Received August 15, 2013; Revised September 29, 2013; Accepted October 12, 2013

ABSTRACT

NPInter (<http://www.bioinfo.org/NPInter>) is a database that integrates experimentally verified functional interactions between noncoding RNAs (excluding tRNAs and rRNAs) and other biomolecules (proteins, RNAs and genomic DNAs). Extensive studies on ncRNA interactions have shown that ncRNAs could act as part of enzymatic or structural complexes, gene regulators or other functional elements. With the development of high-throughput biotechnology, such as cross-linking immunoprecipitation and high-throughput sequencing (CLIP-seq), the number of known ncRNA interactions, especially those formed by protein binding, has grown rapidly in recent years. In this work, we updated NPInter to version 2.0 by collecting ncRNA interactions from recent literature and related databases, expanding the number of entries to 201 107 covering 18 species. In addition, NPInter v2.0 incorporated a service for the BLAST alignment search as well as visualization of interactions.

INTRODUCTION

Interactions of RNA with other biomolecules are fundamental in cellular processes. In particular, noncoding RNA–protein interactions play important roles in protein synthesis (1,2), gene expression (3,4), RNA processing (5–7), developmental regulation (8,9), etc. Interactions between ncRNAs and proteins can be either direct or indirect. Some lncRNAs function as transcriptional regulators through direct association with

transcription factors (10–12), while others indirectly interact with proteins in cooperation with genomic DNAs (13). The functional importance of many ncRNA–protein interactions for correct transcriptional regulation has been demonstrated (14–19), suggesting wide-ranging effects of ncRNA–protein interaction. In addition to targeting or being targeted by proteins or protein-coding transcripts, ncRNAs could also potentially target other ncRNAs, resulting in a layer of regulatory interactions between noncoding RNA classes (11,20). Consequently, cataloguing interactions of ncRNAs and other biomolecules is significant for gaining insight into biological processes and understanding the mechanism by which ncRNAs carry out their regulatory function.

Given the importance of ncRNA interactions in various pathways, we embarked on a project to build a comprehensive catalogue of such data and established the NPInter database (21). A large amount of new research has led to deeper insight into ncRNA interactions at a variety of levels; thus, NPInter has been updated to v2.0 to accommodate for such expanding data resources.

Following data collection as described below, noncoding molecules in NPInter were automatically filtered and assigned identifiers from NONCODE (22) or miRBase (23), while protein-coding molecules were assigned identifiers from UniProt (24), RefSeq (25) or UniGene (26). With the exception of rRNAs and tRNAs, all reported interactions of ncRNAs were included. The aim of the database is to provide a platform that will facilitate both bioinformatic as well as experimental research. In addition to a user-friendly interface and a convenient search option that allow efficient recovery of related interactions and other information, NPInter v2.0 also provides a visualization platform for related interactions. Whole datasets can be

*To whom correspondence should be addressed. Tel: +86 10 6488 8543; Fax: +86 10 6487 7837; Email: crs@sun5.ibp.ac.cn
Correspondence may also be addressed to Yi Zhao. Tel: +86 106 260 1010; Fax: +86 106 260 1356; Email: biozy@ict.ac.cn

The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

downloaded and search results can be exported in text format.

DATA COLLECTION AND ANNOTATION

Based on the first version of NPInter database, new datasets from literature and other related databases were collected. In this update, data from literature accounts for the major source. We first retrieved literature published in the last 5 years from PubMed, employing key words including 'CLIP interaction', 'non-coding RNA bind protein post-transcriptional', 'rna-rna interaction', 'lncRNA protein interaction', 'lncRNA bind', 'RNA protein cross linking', 'RIP non-coding RNA', etc, and found 1270 relevant articles. Information on reported interactions, either verified by experiments or derived from binding sites generated by sequencing, was manually extracted. For the latter, we only extracted processed data by authors, rather than raw sequencing data. Binding sites were first compared with RefSeq coding genes to eliminate those located within coding regions and then screened against NONCODE, which serves as an ncRNA reference database. Binding sites that lie within NONCODE ncRNAs were retained and assigned NONCODE IDs, while others were discarded. As

interaction partners, proteins, protein-coding RNAs and DNA were assigned UniProt IDs, RefSeq IDs and UniGene IDs, respectively, while other noncoding RNAs and DNA were still assigned NONCODE IDs or miRBase IDs. We also integrated data from external resources, mainly LncRNADisease (27), which curated 478 experimentally supported lncRNA interactions at various levels, including binding, regulation and co-expression. Molecules from LncRNADisease were subjected to the same annotation pipeline. By manual curation on data from literature and integration of data from LncRNA Disease, NPInter v2.0 provides a multilevel snapshot of the interactome. A process of redundancy elimination was then performed on the whole dataset, including previously existing data and newly collected data as mentioned above. Redundant interactions were aggregated into a single record. The workflow used for the generation of NPInter is schematically shown in Figure 1.

DATABASE CONTENT AND STRUCTURE

The purpose of the database is to serve as a knowledge base for experimentally-oriented studies and as a resource for bioinformatics applications. The first release of NPInter in 2006 contained 700 published functional

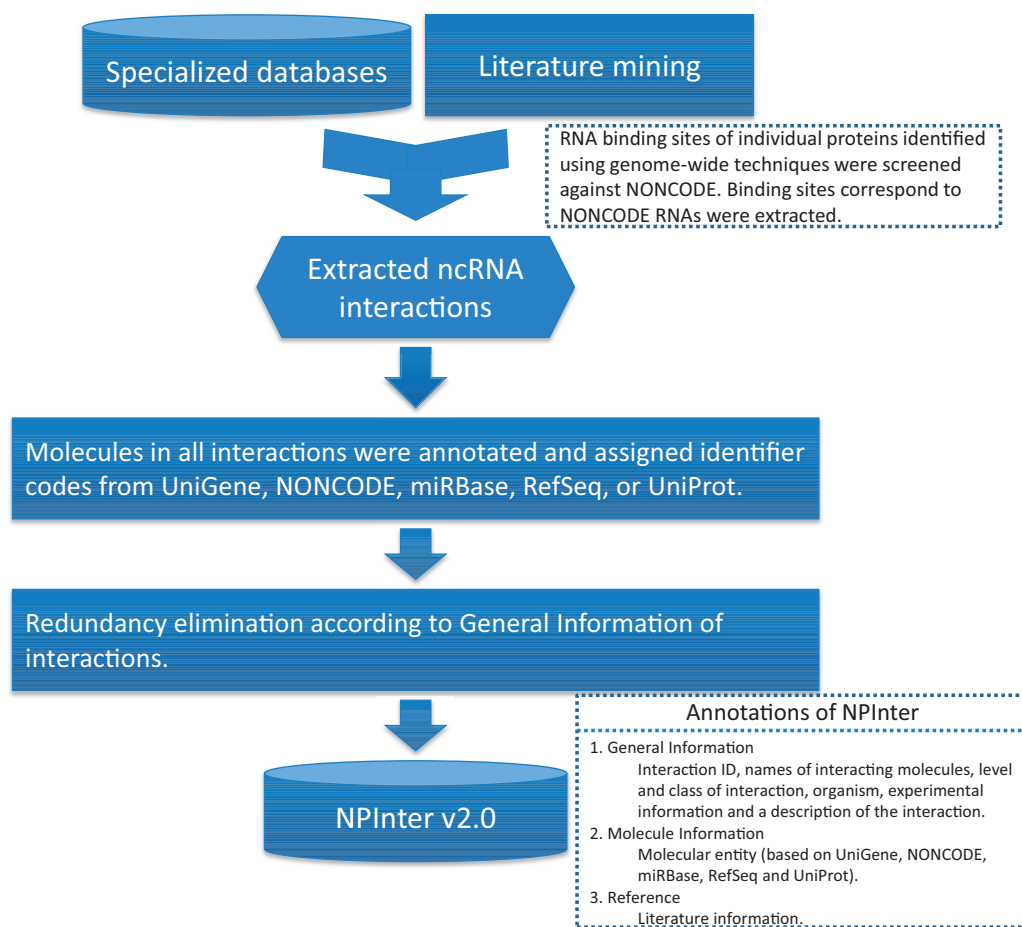


Figure 1. Overview of the NPInter v2.0 Database. NcRNA interactions were obtained from literature and specialized databases. RNA binding sites of individual proteins identified using genome-wide techniques were screened against NONCODE and only those correspond to NONCODE RNAs were extracted.

interactions from six model organisms. NPInter v2.0 presently contains 201 107 ncRNA interactions distributed on 18 organisms, collected from 529 published articles. The significant growth in the amount of data is primarily because of systematic identification of protein binding sites on the transcriptome through a combination of CLIP and RNA sequencing, whereas other interactions, including ncRNA–RNA interactions and transcript factor (TF)–ncRNA interactions, were obtained mainly from interaction studies on individual ncRNA.

The basic information on each interaction contains an interaction ID, name of ncRNA and its interaction partner, organism in which the interaction was identified, level and class of interaction manually defined in NPInter, as well as tags manually added. The interaction partners can be DNA, RNA or protein. For example, at the binding level, which accounts for most of the NPInter datasets, ncRNAs may interact with DNA (13,28), ncRNA (29), miRNA (30), mRNA (31) and protein (32). At the level of indirect interactions, ncRNAs may either regulate or be regulated by an interaction partner (33,34). The level of interaction defined in NPInter represents the types of interacting molecules and characteristics of the interaction, including ‘RNA–protein’, ‘RNA–RNA’ and ‘DNA–TF’. In general, NPInter classifies all

interactions into three classes, which are ‘binding’, ‘regulatory’ and ‘co-expression’. Tags are added to give brief introductions to each interaction, suggesting in which biological process the interaction participates. Our tags are ‘expression correlation’, ‘genetic interaction’, ‘genomic location related’, ‘indirect’, ‘miRNA’, ‘miRNA target interaction’, ‘ncRNA affects synthesis or function of protein’, ‘ncRNA is regulated’, ‘ncRNA–protein binding’, ‘expression, processing, or function of ncRNA is affected’, ‘ncRNA targets mRNA’, ‘other linkages’, ‘promoter as action site’, ‘regulatory’ and ‘RNA–RNA interaction’. Each interaction can be labelled simultaneously with several tags. (See Supplementary Data: Supplementary I).

The NPInter database is composed of three linked tables:

- The Interaction table gives a description of details of the interaction between two molecules. For example, for the interaction between Kcnqlot1 and Dnmt1 in mouse, information in this table states that this interaction occurs at the level of ‘RNA–protein’ and is classified into ‘binding’ class with a tag ‘ncRNA–protein binding’.
- The Molecule table describes information on interacting molecules, containing identifiers from UniGene,

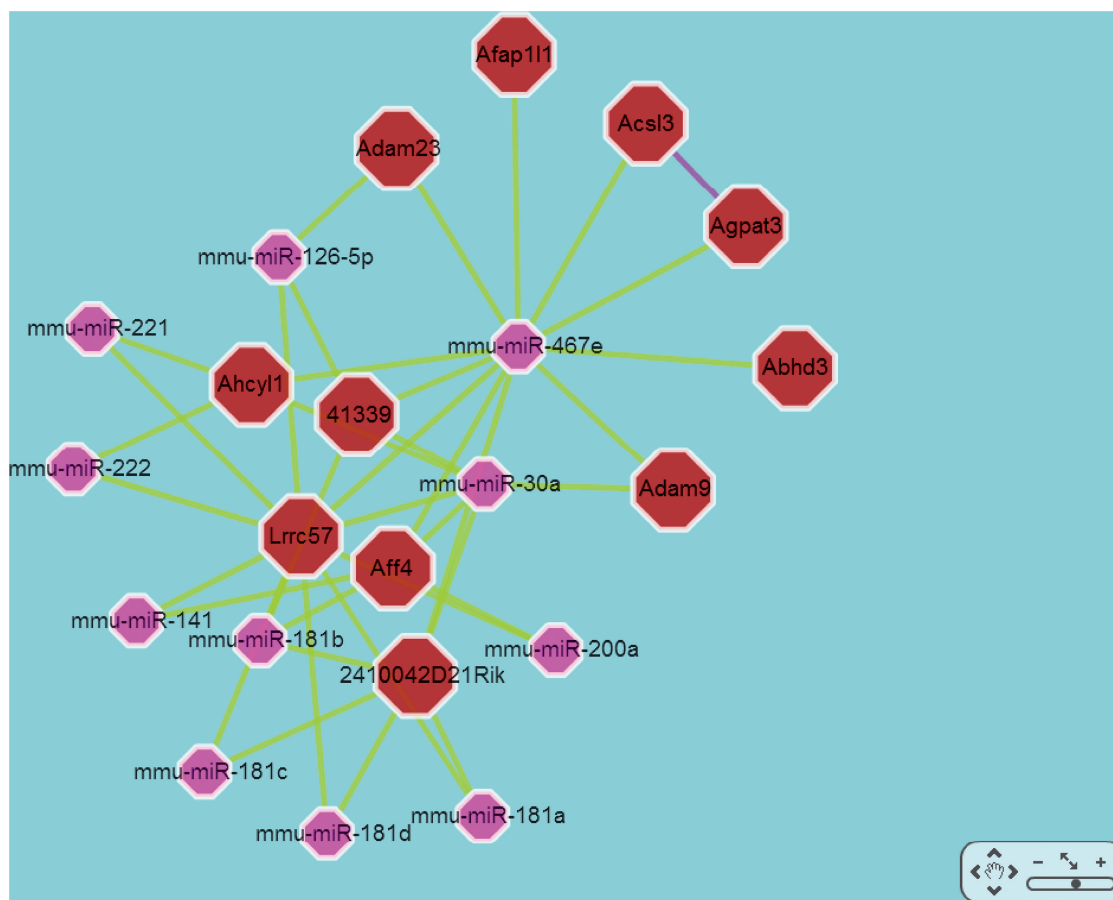


Figure 2. Graph visualization of NPInter interactions. Node colour and size indicates molecule type, while edge colour indicates interaction source. The initial placement of the nodes is determined by a ‘force-directed’ layout algorithm that aims to keep the more similar nodes closer together, but the placement may later be adjusted by the user.

NONCODE, miRBase, RefSeq or UniProt as well as the name, aliases and description of each molecule in NPInter.

- The Reference table gives the details of literature references in the interaction table. Each record in the table includes the MEDLINE standard article code (PMID) as well as general publication information.

DATA ACCESS AND VISUALIZATION

NPInter allows users to browse interactions by species, ncRNA classes or interaction tags. Users can also query the database through the Search interface, using the name or aliases of molecules, molecule IDs or any other descriptive words. The whole datasets of NPInter can be directly downloaded from the webpage, and the results of each search can be exported. In the updated version of the database, we have also integrated the online BLAST service (NCBI www.BLAST version 2.2.17), which allows sequence similarity searches for both nucleotide and peptide to be run on NPInter entries. Importantly, for a noncoding RNA with no recognized name or identified function, it is also possible to search its potential interactions with other molecules in NPInter simply based on its sequence. Additionally, NPInter also offers a graph-based visualization of interactions (Figure 2). In the visualization graph, each molecule is a node and interactions between molecules are designated as edges. Red nodes represent proteins or protein-coding RNAs and DNA, while blue and pink nodes represent ncRNAs and miRNAs, respectively. Information of each molecule can be checked in a new window by clicking on the corresponding node. The colour of the edges discriminates the source of each interaction, green suggesting the interaction is from NPInter and purple suggesting the interaction is from STRING (35). The location of the nodes is determined by the ForceDirected layout algorithm as implemented in Cytoscape Web (36). The graph can be opened in the stand-alone webpage to produce high resolution images in which size, colour and location of nodes and edges can be adjusted. To maintain an up-to-date and comprehensive resource, we encourage users to submit newly published ncRNA interactions, with PubMed accession numbers required.

CONCLUSION

Noncoding RNAs have emerged as key molecular players in different biological processes (37). Characterizing ncRNA interactions will largely contribute towards the discovery of novel ncRNA functions. However, existing databases are not enough to provide a comprehensive resource for such a purpose. For example, starBase (38) lists all binding sites of several proteins in the transcriptome, but there is no specific annotation for noncoding transcripts. For another example, NPIDB (39) focuses on nucleic acid–protein complexes from PDB, thus, most RNAs included are either rRNAs or tRNAs.

In contrast, NPInter is more informative on ncRNA interactions than other databases, with aims on

integrating interactions of the other types of ncRNAs that serve a variety of functions. In other words, NPInter is one of the most comprehensive databases of interactions between ncRNAs and other biomolecules.

Compared with the previous version of NPInter, the new version is a step towards a more integrated knowledge database. The total number of functional interactions of ncRNAs has been expanded. NPInter v2.0 is also developed to present graphical molecular interaction networks that will enable biological scientists to explore their data in a more systems-oriented manner. NPInter will continue to keep track of and promptly collect new interactions. The authors plan to update NPInter whenever there is an accumulation of novel ncRNA interactions reported in literature or other sources.

It is worth mentioning that NPInter is included in our systematic platform for noncoding RNAs. Our platform consists of ncRNA resources such as NONCODE and NPInter, as well as online tools and web servers (40–43) including CNCI (40) and ncFANs (43) for analysis of ncRNAs. As a member of the union of specific databases and tools for noncoding RNAs, NPInter is expected to remain an informative and valuable data source on the biological roles of ncRNAs for the scientific community.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Andrew Plygawko and Jianjun Luo for carefully reading our manuscript.

FUNDING

Funding for open access charge: National High-tech Research and Development Projects 863 [2012AA020402]; Training Program of the Major Research plan of the National Natural Science Foundation of China [91229120]; National Natural Science Foundation of China [31000586]; Chinese Academy of Science Strategic Project of Leading Science and Technology [XDA01020402]; National Key Basic Research and Development Program 973 [2009CB825401]; the National Center for Mathematics and Interdisciplinary Sciences.

Conflict of interest statement. None declared.

REFERENCES

1. Moore, P.B. (1998) The three-dimensional structure of the ribosome and its components. *Annu. Rev. Biophys. Biomol. Struct.*, **27**, 35–58.
2. Ramakrishnan, V. and White, S.W. (1998) Ribosomal protein structures: insights into the architecture, machinery and evolution of the ribosome. *Trends Biochem. Sci.*, **23**, 208–212.
3. Siomi, H. and Dreyfuss, G. (1997) RNA-binding proteins as regulators of gene expression. *Curr. Opin. Genet. Dev.*, **7**, 345–353.
4. Mata, J., Marguerat, S. and Bahler, J. (2005) Post-transcriptional control of gene expression: a genome-wide perspective. *Trends Biochem. Sci.*, **30**, 506–514.

5. Singh, R. (2002) RNA-protein interactions that regulate pre-mRNA splicing. *Gene Expr.*, **10**, 79–92.
6. Varani, G. and Nagai, K. (1998) RNA recognition by RNP proteins during RNA processing. *Annu. Rev. Biophys. Biomol. Struct.*, **27**, 407–445.
7. Frank, D.N. and Pace, N.R. (1998) Ribonuclease P: unity and diversity in a tRNA processing ribozyme. *Annu. Rev. Biochem.*, **67**, 153–180.
8. Hall, K.B. (2002) RNA-protein interactions. *Curr. Opin. Struct. Biol.*, **12**, 283–288.
9. Tian, B., Bevilacqua, P.C., Diegelman-Parente, A. and Mathews, M.B. (2004) The double-stranded-RNA-binding motif: interference and much more. *Nat. Rev. Mol. Cell Biol.*, **5**, 1013–1023.
10. Mercer, T.R., Dinger, M.E. and Mattick, J.S. (2009) Long non-coding RNAs: insights into functions. *Nat. Rev. Genet.*, **10**, 155–159.
11. Wang, K.C. and Chang, H.Y. (2011) Molecular mechanisms of long noncoding RNAs. *Mol. Cell*, **43**, 904–914.
12. Guttman, M. and Rinn, J.L. (2012) Modular regulatory principles of large non-coding RNAs. *Nature*, **482**, 339–346.
13. Schmitz, K.M., Mayer, C., Postepska, A. and Grummt, I. (2010) Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes. *Genes Dev.*, **24**, 2264–2269.
14. Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Bruggmann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E. *et al.* (2007) Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell*, **129**, 1311–1323.
15. Guttman, M., Donaghey, J., Carey, B.W., Garber, M., Grenier, J.K., Munson, G., Young, G., Lucas, A.B., Ach, R., Bruhn, L. *et al.* (2011) lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature*, **477**, 295–300.
16. Khalil, A.M., Guttman, M., Huarte, M., Garber, M., Raj, A., Rivea Morales, D., Thomas, K., Presser, A., Bernstein, B.E., van Oudenarden, A. *et al.* (2009) Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc. Natl Acad. Sci. USA*, **106**, 11667–11672.
17. Maison, C., Bailly, D., Peters, A.H., Quivy, J.P., Roche, D., Taddei, A., Lachner, M., Jenuwein, T. and Almouzni, G. (2002) Higher-order structure in pericentric heterochromatin involves a distinct pattern of histone modification and an RNA component. *Nat. Genet.*, **30**, 329–334.
18. Bernstein, E., Duncan, E.M., Masui, O., Gil, J., Heard, E. and Allis, C.D. (2006) Mouse polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin. *Mol. Cell Biol.*, **26**, 2560–2569.
19. Wutz, A., Rasmussen, T.P. and Jaenisch, R. (2002) Chromosomal silencing and localization are mediated by different domains of Xist RNA. *Nat. Genet.*, **30**, 167–174.
20. Jalali, S., Bhartiya, D., Lalwani, M.K., Sivasubbu, S. and Scaria, V. (2013) Systematic transcriptome wide analysis of lincRNA-miRNA interactions. *PLoS One*, **8**, e53823.
21. Wu, T., Wang, J., Liu, C., Zhang, Y., Shi, B., Zhu, X., Zhang, Z., Skogerbo, G., Chen, L., Lu, H. *et al.* (2006) NPInter: the noncoding RNAs and protein related biomacromolecules interaction database. *Nucleic Acids Res.*, **34**, D150–D152.
22. Bu, D., Yu, K., Sun, S., Xie, C., Skogerbo, G., Miao, R., Xiao, H., Liao, Q., Luo, H., Zhao, G. *et al.* (2012) NONCODE v3.0: integrative annotation of long noncoding RNAs. *Nucleic Acids Res.*, **40**, D210–D215.
23. Kozomara, A. and Griffiths-Jones, S. (2011) miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res.*, **39**, D152–D157.
24. UniProt Consortium. (2013) Update on activities at the Universal Protein Resource (UniProt) in 2013. *Nucleic Acids Res.*, **41**, D43–D47.
25. Pruitt, K.D., Tatusova, T., Brown, G.R. and Maglott, D.R. (2012) NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. *Nucleic Acids Res.*, **40**, D130–D135.
26. NCBI Resource Coordinators. (2013) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **41**, D8–D20.
27. Chen, G., Wang, Z., Wang, D., Qiu, C., Liu, M., Chen, X., Zhang, Q., Yan, G. and Cui, Q. (2013) LncRNADisease: a database for long-non-coding RNA-associated diseases. *Nucleic Acids Res.*, **41**, D983–D986.
28. Schorderet, P. and Duboule, D. (2011) Structural and functional differences in the long non-coding RNA hotair in mouse and human. *PLoS Genet.*, **7**, e1002071.
29. Smaldone, G.T., Revelles, O., Gaballa, A., Sauer, U., Antelmann, H. and Helmman, J.D. (2012) A global investigation of the *Bacillus subtilis* iron-sparing response identifies major changes in metabolism. *J. Bacteriol.*, **194**, 2594–2605.
30. Clark, M.B. and Mattick, J.S. (2011) Long noncoding RNAs in cell biology. *Semin. Cell Dev. Biol.*, **22**, 366–376.
31. Lipovich, L., Johnson, R. and Lin, C.Y. (2010) MacroRNA underdogs in a microRNA world: evolutionary, regulatory, and biomedical significance of mammalian long non-protein-coding RNA. *Biochim. Biophys. Acta*, **1799**, 597–615.
32. Wood, H., Lührink, J. and Tollervey, D. (1992) Evolutionary conserved nucleotides within the *E. coli* 4.5S RNA are required for association with P48 *in vitro* and for optimal function *in vivo*. *Nucleic Acids Res.*, **20**, 5919–5925.
33. Wang, Z. and Roeder, R.G. (1995) Structure and function of a human transcription factor TFIIB subunit that is evolutionarily conserved and contains both TFIIB- and high-mobility-group protein 2-related domains. *Proc. Natl Acad. Sci. USA*, **92**, 7026–7030.
34. Luo, Y., Kurz, J., MacAfee, N. and Krause, M.O. (1997) C-myc deregulation during transformation induction: involvement of 7SK RNA. *J. Cell Biochem.*, **64**, 313–327.
35. Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., Lin, J., Minguez, P., Bork, P., von Mering, C. *et al.* (2013) STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.*, **41**, D808–D815.
36. Saito, R., Smoot, M.E., Ono, K., Ruschinski, J., Wang, P.L., Lotia, S., Pico, A.R., Bader, G.D. and Ideker, T. (2012) A travel guide to Cytoscape plugins. *Nat. Methods*, **9**, 1069–1076.
37. Beckedorff, F.C., Amaral, M.S., Deocesano-Pereira, C. and Verjovski-Almeida, S. (2013) Long noncoding RNAs and their implications in cancer epigenetics. *Biosci. Rep.*, **33**, pii: e00061.
38. Yang, J.H., Li, J.H., Shao, P., Zhou, H., Chen, Y.Q. and Qu, L.H. (2011) starBase: a database for exploring microRNA-mRNA interaction maps from Argonaute CLIP-Seq and Degradome-Seq data. *Nucleic Acids Res.*, **39**, D202–D209.
39. Kirsanov, D.D., Zanevina, O.N., Aksianov, E.A., Spirin, S.A., Karyagina, A.S. and Alexeevski, A.V. (2013) NPIDB: Nucleic acid-Protein Interaction DataBase. *Nucleic Acids Res.*, **41**, D517–D523.
40. Sun, L., Luo, H., Bu, D., Zhao, G., Yu, K., Zhang, C., Liu, Y., Chen, R. and Zhao, Y. (2013) Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Res.*, **41**, e166.
41. Liao, Q., Liu, C., Yuan, X., Kang, S., Miao, R., Xiao, H., Zhao, G., Luo, H., Bu, D., Zhao, H. *et al.* (2011) Large-scale prediction of long non-coding RNA functions in a coding-non-coding gene co-expression network. *Nucleic Acids Res.*, **39**, 3864–3878.
42. Guo, X., Gao, L., Liao, Q., Xiao, H., Ma, X., Yang, X., Luo, H., Zhao, G., Bu, D., Jiao, F. *et al.* (2013) Long non-coding RNAs function annotation: a global prediction method based on bi-colored networks. *Nucleic Acids Res.*, **41**, e35.
43. Liao, Q., Xiao, H., Bu, D., Xie, C., Miao, R., Luo, H., Zhao, G., Yu, K., Zhao, H., Skogerbo, G. *et al.* (2011) ncFANs: a web server for functional annotation of long non-coding RNAs. *Nucleic Acids Res.*, **39**, W118–W124.